



## The human secretome

Downloaded from: <https://research.chalmers.se>, 2023-05-05 05:28 UTC

Citation for the original published paper (version of record):

Uhlen, M., Karlsson, M., Hober, A. et al (2019). The human secretome. Science signaling, 12(609).  
<http://dx.doi.org/10.1126/scisignal.aaz0274>

N.B. When citing this work, cite the original published paper.

## BIOCHEMISTRY

## The human secretome

Mathias Uhlén<sup>1,2,3\*</sup>, Max J. Karlsson<sup>1</sup>, Andreas Hober<sup>1</sup>, Anne-Sophie Svensson<sup>4</sup>, Julia Scheffel<sup>4</sup>, David Kotol<sup>1</sup>, Wen Zhong<sup>1</sup>, Abdellah Tebani<sup>1</sup>, Linnéa Strandberg<sup>1</sup>, Fredrik Edfors<sup>1,5</sup>, Evelina Sjöstedt<sup>3</sup>, Jan Mulder<sup>3</sup>, Adil Mardinoglu<sup>1</sup>, Anna Berling<sup>4</sup>, Siri Ekblad<sup>4</sup>, Melanie Dannemeyer<sup>4</sup>, Sara Kanje<sup>4</sup>, Johan Rockberg<sup>4</sup>, Magnus Lundqvist<sup>4</sup>, Magdalena Malm<sup>4</sup>, Anna-Luisa Volk<sup>4</sup>, Peter Nilsson<sup>1</sup>, Anna Månberg<sup>1</sup>, Tea Dodig-Crnkovic<sup>1</sup>, Elisa Pin<sup>1</sup>, Martin Zwahlen<sup>1</sup>, Per Oksvold<sup>1</sup>, Kalle von Feilitzen<sup>1</sup>, Ragna S. Häussler<sup>1</sup>, Mun-Gwan Hong<sup>1</sup>, Cecilia Lindskog<sup>6</sup>, Fredrik Ponten<sup>6</sup>, Borbala Katona<sup>6</sup>, Jimmy Vuu<sup>6</sup>, Emil Lindström<sup>6</sup>, Jens Nielsen<sup>7</sup>, Jonathan Robinson<sup>7</sup>, Burcu Ayoglu<sup>1</sup>, Diana Mahdessian<sup>1</sup>, Devin Sullivan<sup>1</sup>, Peter Thul<sup>1</sup>, Frida Danielsson<sup>1</sup>, Charlotte Stadler<sup>1</sup>, Emma Lundberg<sup>1</sup>, Göran Bergström<sup>8,9</sup>, Anders Gummesson<sup>8</sup>, Bjørn G. Voldborg<sup>2</sup>, Hanna Tegel<sup>4</sup>, Sophia Hober<sup>4</sup>, Björn Forsström<sup>1</sup>, Jochen M. Schwenk<sup>1</sup>, Linn Fagerberg<sup>1</sup>, Åsa Sivertsson<sup>1</sup>

The proteins secreted by human cells (collectively referred to as the secretome) are important not only for the basic understanding of human biology but also for the identification of potential targets for future diagnostics and therapies. Here, we present a comprehensive analysis of proteins predicted to be secreted in human cells, which provides information about their final localization in the human body, including the proteins actively secreted to peripheral blood. The analysis suggests that a large number of the proteins of the secretome are not secreted out of the cell, but instead are retained intracellularly, whereas another large group of proteins were identified that are predicted to be retained locally at the tissue of expression and not secreted into the blood. Proteins detected in the human blood by mass spectrometry-based proteomics and antibody-based immunoassays are also presented with estimates of their concentrations in the blood. The results are presented in an updated version 19 of the Human Protein Atlas in which each gene encoding a secretome protein is annotated to provide an open-access knowledge resource of the human secretome, including body-wide expression data, spatial localization data down to the single-cell and subcellular levels, and data about the presence of proteins that are detectable in the blood.

## INTRODUCTION

An important class of human proteins are those that are actively transported within the secretory pathway for destinations outside the cytoplasm and nucleus of the cell. The collection of actively secreted proteins, herein referred to as the “human secretome,” constitutes a large fraction of the targets for pharmaceutical drugs, but these are also important as diagnostic targets both for classical clinical chemistry and as potential targets for future precision medicine efforts (1, 2). Many of these proteins are also involved in signaling functions both locally and systemically, including proteins such as cytokines, growth factors, and hormones. Despite the huge interest in this class of proteins, there have been few attempts to define the size and constituents of the entire human secretome. A prediction of the number of putatively secreted proteins, defined as having a signal sequence and no transmembrane regions, was previously estimated to correspond to 2918 protein-coding genes, thus involving

approximately 15% of all human genes (3). Attempts to characterize the proteins present in blood (the “plasma proteome”) have led to the development of multiplex assays involving thousands of protein targets using nucleic acid-based technologies (4, 5), immune-based assays (6), or mass spectrometry (MS) (7, 8). However, these plasma proteome efforts do not normally distinguish between actively secreted proteins (here defined as being part of the secretome) and proteins that are leaked by the millions of cells undergoing cell death at any given moment. In addition, many of the proteins secreted from human cells are not destined for the peripheral blood.

We therefore decided to stratify the actively secreted proteins in humans to define the spatial distribution of each protein with regard to its origin of expression and to provide a comprehensive list of annotated proteins based on their final localization in the body. Starting with a bioinformatics definition of the secretome, the proteins were classified into three major categories: (i) the blood proteins, (ii) the locally secreted proteins, and (iii) the intracellular proteins. The latter might sound counterintuitive, but this category reflects the fact that many proteins secreted into the endoplasmic reticulum (ER) are sorted to various intracellular compartments, such as mitochondria and lysosomes, or they are even retained in the ER or Golgi. These proteins are thus not secreted out of the cell. The locally secreted proteins can further be classified into the various sublocalizations, such as the brain, as well as male and female tissues, and they also include proteins secreted to the digestive tract or those that end up in the extracellular matrix. Last, we annotated the proteins predicted to end up in peripheral blood, thus having important effects in the

<sup>1</sup>Department of Protein Science, Science for Life Laboratory, KTH–Royal Institute of Technology, Stockholm, Sweden. <sup>2</sup>Center for Biosustainability, Technical University of Denmark, Lyngby, Denmark. <sup>3</sup>Department of Neuroscience, Karolinska Institute, Stockholm, Sweden. <sup>4</sup>Department of Protein Science, AlbaNova University Center, KTH–Royal Institute of Technology, Stockholm, Sweden. <sup>5</sup>Department of Genetics, School of Medicine, Stanford University, Stanford, CA, USA. <sup>6</sup>Department of Pathology, Uppsala University, Uppsala, Sweden. <sup>7</sup>Department of Chemical and Biological Engineering, Chalmers University of Technology, Gothenburg, Sweden. <sup>8</sup>Department of Molecular and Clinical Medicine, Institute of Medicine, Sahlgrenska Academy, University of Gothenburg, Gothenburg, Sweden. <sup>9</sup>Region Västra Götaland, Sahlgrenska University Hospital, Department of Clinical Physiology, Gothenburg, Sweden.

\*Corresponding author. Email: mathias.uhlen@scilifelab.se

human body, contributing to the systems-level control of homeostasis, transport of nutrients, inflammatory response, defense mechanisms, hormone regulation, and many other functions. To complement this annotation of the human secretome, we also investigated the proteins found in blood using various technology platforms. All results are presented in a new version of the Human Protein Atlas (HPA) ([www.proteinatlas.org/blood](http://www.proteinatlas.org/blood)) with a “secretome” part including data on each gene encoding these proteins.

## RESULTS

### Revised list of the human secretome

We first performed a protein-centric transcriptomics scan to define a revised set of human secreted proteins (secretome) based on 19,670 protein-coding genes predicted by Ensembl (9). For each protein-coding gene, all protein isoforms (splice variants) were annotated (Fig. 1A) on the basis of the presence of a signal peptide, transmembrane regions, or both, and each protein isoform was classified as being secreted, membrane bound, or intracellular. The secreted proteins are here defined according to the HPA classification (3) as those proteins that have a signal peptide but lack a transmembrane region. The 338 Ensembl immunoglobulin genes (file S1) were excluded because of their complex genetic structure, which consists of many partial genes with variable (V), diversity (D), and joining (J) regions. In addition, proteins annotated as being secreted by UniProt (10) were added, which resulted in a list of 3513 genes with at least one predicted secreted isoform (transcript). From this list, proteins with no corresponding entry in the current Ensembl version (v92) and genes for which the predicted secreted isoform has low evidence for existence were excluded. The revised number of genes encoding potentially secreted proteins (the human secretome) was 2641 (file S2). This list comprises approximately 13% of all human protein-coding genes, and it serves as a resource for all researchers interested in secreted proteins as targets for diagnostic and therapeutic drugs.

### Annotation of the human secretome into localization classes

On the basis of the published literature, bioinformatics analysis, and experimental evidence, all genes coding for the putative secretome were subsequently manually annotated, taking into account their spatial distribution in the human body. Starting with a bioinformatics definition of the secretome, the proteins were classified into three major categories: (i) the blood proteins, (ii) the locally secreted proteins, and (iii) the intracellular proteins. For the latter category, bioinformatics evidence for intracellular locations was also used, such as ER retention signals (11), a peroxisome-targeting signal (12), and a mitochondrial target signal (13). The locally secreted proteins were further subdivided into seven classes to yield altogether nine protein categories (Fig. 1A). Note that the annotation considers the possible functional role of each protein and, as an example, the well-known plasma protein PSA (*KLK3*) was here annotated as being localized to male reproductive tissue and not the blood, because it was here assumed that the primary functional role of this protein is in the prostate. The annotation of all 2641 protein-coding genes identified 730 proteins as blood secretome proteins, and about 500 were annotated as being secreted to local compartments, that is, male or female reproductive tissues, the brain, or other tissues, such as the eye or the skin. Eighty-eight proteins were identified as being secreted to the gastrointestinal tract, including 30 proteins produced in the pancreas and 25 in the salivary glands (fig. S1). More than

200 proteins were further annotated to belong to the group of proteins involved in the forming and function of the extracellular matrix, including both structural proteins, such as laminins, collagens, elastin, and fibronectin, and the matricellular proteins (14). More than 900 proteins were annotated to be intracellular or membrane related, with 254 of these proteins predicted to be localized to the Golgi, ER, or both and 270 proteins being membrane associated. Last, about 170 genes encoded proteins that lacked supporting data for their location. These are again interesting proteins for further studies to explore their function and location. A list of the genes in each category was compiled (file S2) and is also available at the open-access Blood Atlas resource ([www.proteinatlas.org/blood](http://www.proteinatlas.org/blood)). Given that 932 genes were annotated to encode intracellularly localized proteins, despite having a predicted secreted isoform, the genome-wide annotation of the subcellular location of the protein-coding genes was revised (Fig. 1B). The resulting list of the human secretome consists of 1709 genes with at least one isoform (transcript) coding for a predicted secreted protein. Note that many genes encode multiple transcripts predicted to be in more than one location.

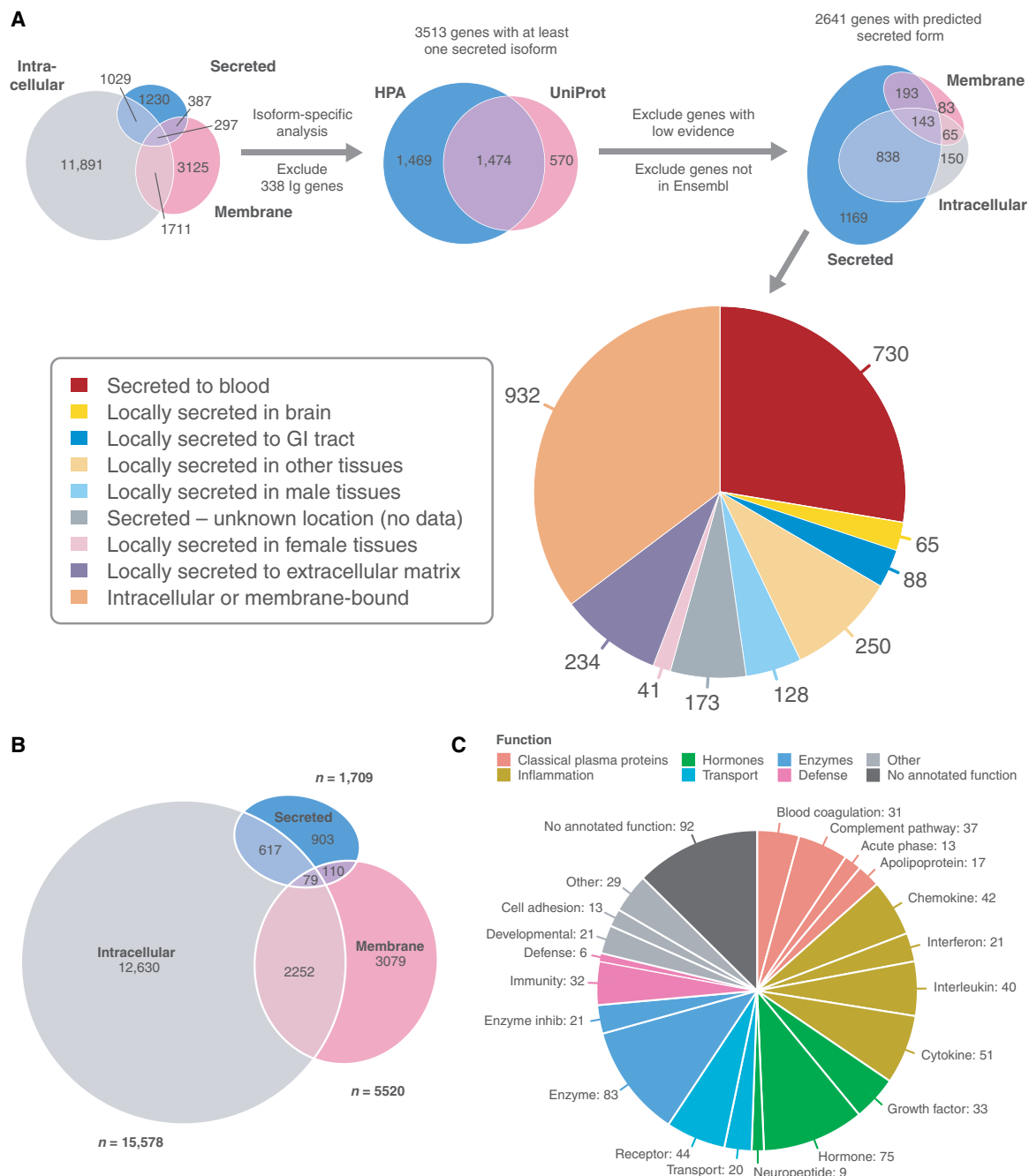
### Functional analysis of those proteins predicted to be actively secreted to the blood

A functional analysis (Fig. 1C) of the 730 proteins predicted to be secreted to human blood revealed many well-characterized proteins, such as cytokines, interleukins, interferons, and chemokines (altogether 154 proteins), complement and coagulation factors ( $n = 68$  proteins), hormones ( $n = 75$ ), growth factors ( $n = 33$ ), and enzymes ( $n = 83$ ). Many of these proteins are interesting pharmaceutical targets, and 72 are already the products or targets of U.S. Food and Drug Administration (FDA)-approved drugs (file S3) (15). Furthermore, almost 100 of the predicted secreted blood proteins have currently no functional annotation in UniProt; thus, these are interesting targets for further exploration of their functional role in the blood.

### The tissue distribution of the human secretome proteins

The tissue expression pattern of the human secretome genes was subsequently analyzed on the basis of transcriptomics data from the HPA. A classification according to tissue specificity as previously described (3) was performed for all of the secretome genes in the different annotation categories. The results showed that genes predicted to be locally expressed mainly consisted of tissue-enriched genes. Furthermore, genes encoding intracellular proteins generally showed low tissue specificity (Fig. 2A), which suggests that they have a more “house-keeping” role in the cell. Similarly, the matrix proteins are widely expressed across the analyzed tissues. A large portion of the blood proteins are produced in the liver, whereas others are generated either by blood cells or more generally across all tissues.

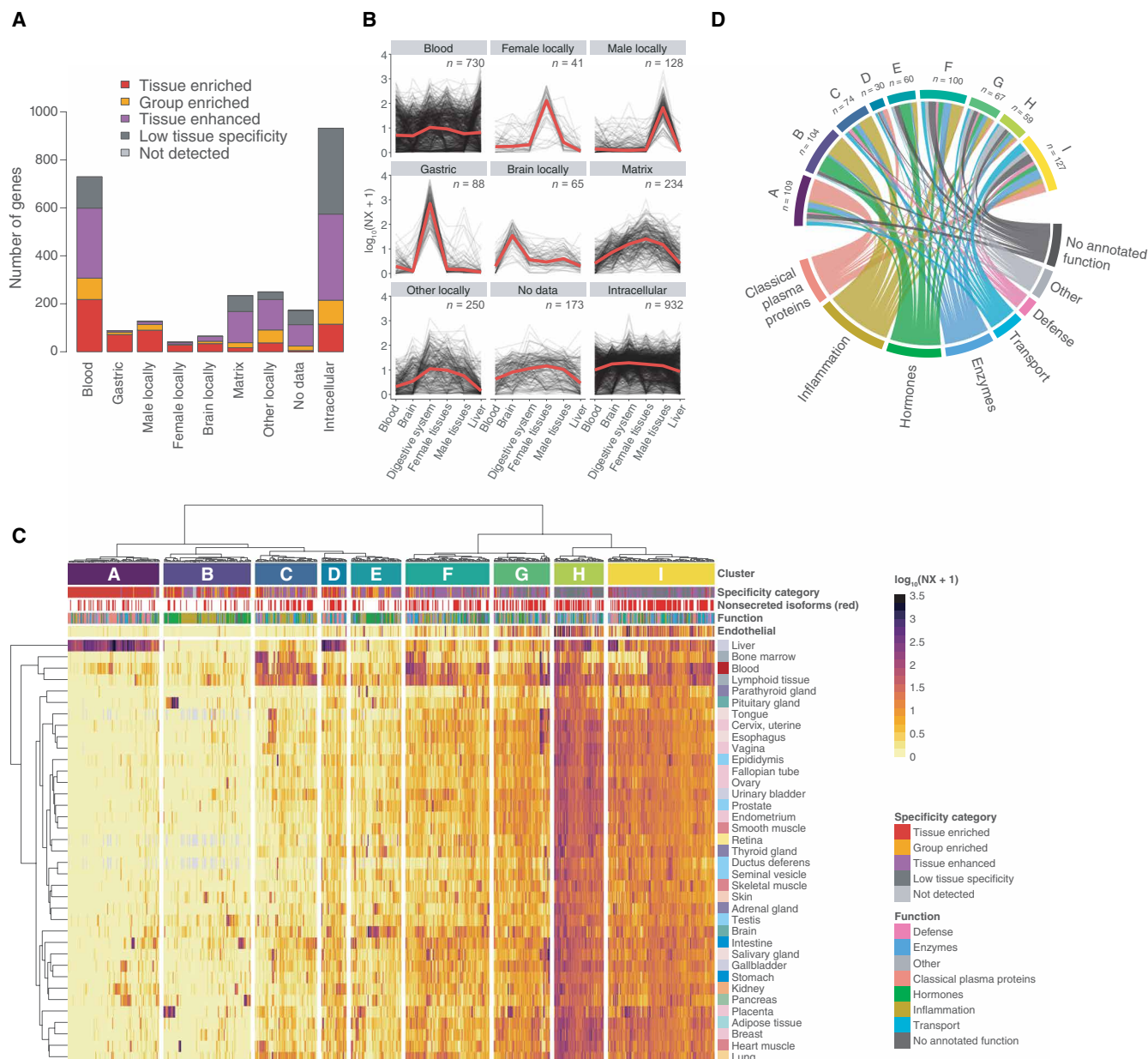
A cluster analysis (16) suggests that the expression profiles across the 730 blood proteins can be stratified into nine expression clusters (Fig. 2, B and C), and these can further be grouped according to tissue origin, including 139 proteins mainly originating from the liver (clusters A and D) with classical plasma proteins, such as albumin, transferrin, the apolipoproteins, and complement factors. Another group consists of the 174 proteins mainly originating from human blood cells, lymphoid tissues, or both (clusters C and F), which include many chemokines and granzymes. The 60 proteins that are mainly enriched in the brain (cluster E) constitute a third group to which several hormones and neuropeptides belong, including oxytocin, gonadotropin-releasing hormone, and pro-melanin



**Fig. 1. Annotation of the human secretome.** (A) Workflow used to predict the final location of components of the human secretome ( $n = 2641$  proteins). The proteins are classified according to their predicted final location as seen in the pie chart (see file S2 for a complete list). (B) Predicted number of genes in the three indicated categories after considering the intracellular and membrane-bound annotations. (C) Functional analysis of the 730 proteins annotated as being blood proteins. The color code represents eight functional classes, including the subclasses denoted in the pie chart.

concentrating hormone. A fourth group includes 104 proteins with selective expression (cluster B), including interferons and hormones produced in specialized tissues, such as the placenta or endocrine glands. Last, there is a large group of 253 proteins that are expressed ubiquitously across many tissues (clusters G, H, and I). Many of these proteins are highly expressed in endothelial cells present throughout the body, which could explain their diverse expression, but it is also possible that some of the expression is related to the

nonsecreted isoforms of the respective gene products and more in-depth studies are needed to resolve the isoform-specific transcript abundances for these genes. The relationship between the clusters and protein functions are shown as a chord diagram (Fig. 2D), demonstrating that the genes in cluster B mainly encode cytokines and growth factors, whereas genes in cluster A encode not only complement, coagulation factors, and acute phase proteins but also transport proteins and enzymes.



**Fig. 2. The tissue distribution of the human secretome proteins.** (A) Tissue specificity classification based on transcriptomics data for the genes in the indicated annotation categories. (B) Expression values ( $\log_2(NX + 1)$ ) across six selected tissues (see Fig. S2 for the corresponding heatmaps) for the nine annotation categories (A to H). The number of protein-coding genes in each category is shown (n). (C) Heatmap based on the relative expression of all blood secretome protein-coding genes (n = 730) with nine expression clusters shown on top. (D) Chord diagram showing the relationship between the functional annotation of the 730 blood secretome proteins (bottom) with the nine expression clusters (top).

### Proteins detected in blood by MS

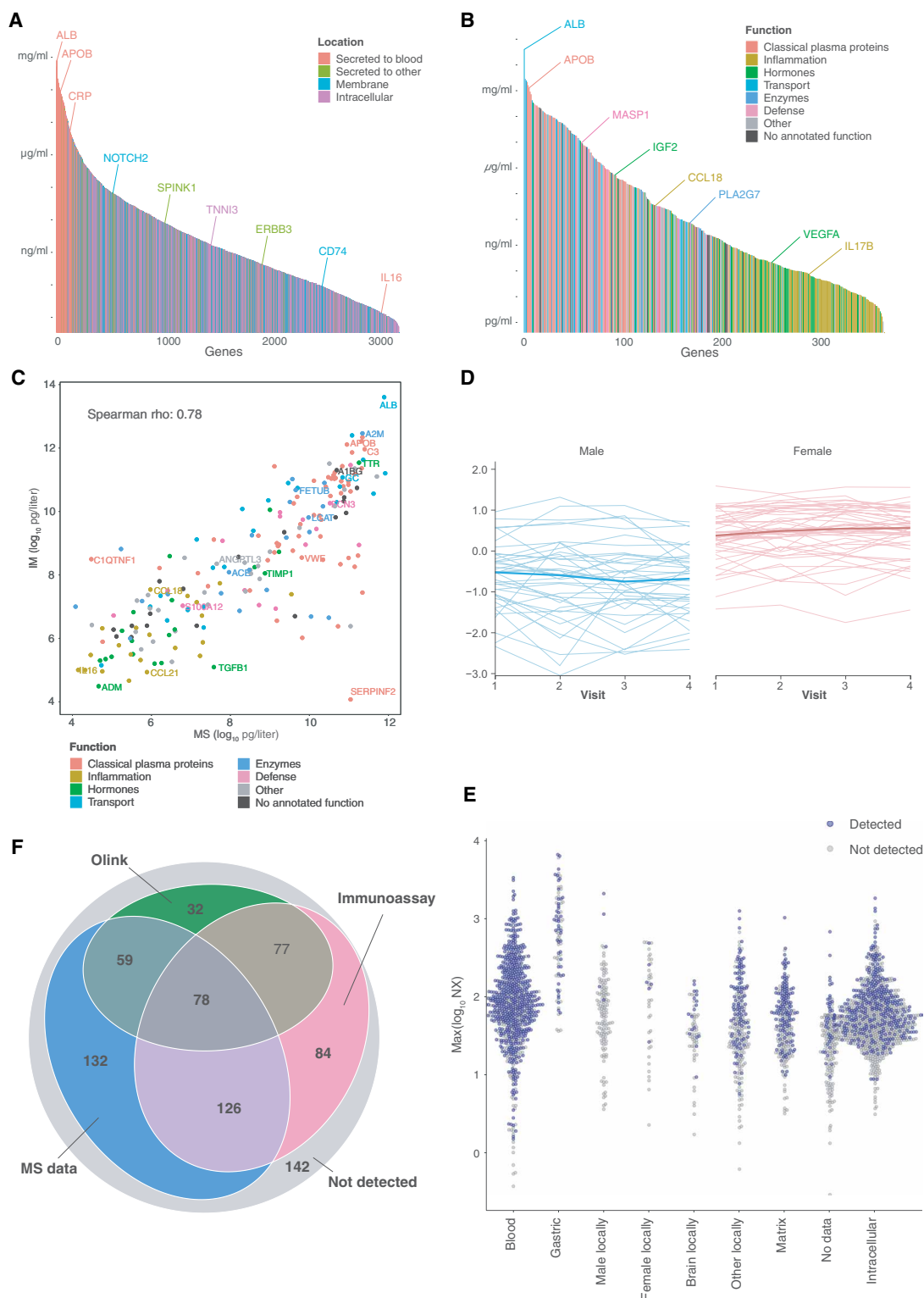
We subsequently decided to investigate the presence of the predicted secretome proteins in human blood based on MS and antibody-based immunoassays and to complement this with “in-house” analysis using a sensitive proximity extension assay (PEA) (17). For MS-based proteomics, protein concentrations in blood (Fig. 3A) were inferred by spectral counting (18) from a combined dataset of more than 170 publicly available experiments hosted in the Human Plasma PeptideAtlas (19). More than 3000 proteins were detected in this diverse set of experiments (file S4). The detected proteins were colored according to their predicted localization,

with proteins annotated to be intracellular, membrane bound, or secreted and with the latter group stratified into those secreted to blood and those secreted to other compartments (according to Fig. 1). In this regard, note that there are a few abundant proteins that make up most of the protein mass in blood, with 99% of the protein mass corresponding to only a few proteins (20). However, many proteins are present at lower concentrations either resulting from active secretion or proteins leaked from cells due to normal cell turnover or pathology involving various diseases. The analysis showed that most of the proteins detected by MS were here defined as leakage proteins.



**Fig. 3. Proteins detected in human blood.**

The predicted amounts of proteins detected in human blood by three different technology platforms. **(A)** Plasma concentrations based on MS assays for proteins detected in blood. The plot shows plasma concentrations estimated from spectral counting of 3223 proteins belonging to MS experiments in the Human Plasma PeptideAtlas. The bars are colored on the basis of the classification of the corresponding proteins into one of four localization categories: secreted to blood, secreted to other, membrane bound, or intracellular. Some examples of proteins are shown as reference. **(B)** Plasma concentrations based on immunoassays from reference articles for 365 proteins actively secreted to blood. The bar plot shows ranked median plasma concentrations with bars colored on the basis of the classification of the corresponding proteins into one of the eight indicated functional categories. **(C)** Plot showing the correlation of the estimated concentrations of 205 blood secretome proteins with data from both proteomics and immunoassays. **(D)** The protein profiles of 86 individuals during a 1-year period were analyzed using a set of multiplex PEAs targeting 748 proteins. The image shows an example of the average amounts of a protein (leptin) for four visits spanning 1 year. The individuals were stratified according to females (red) and males (blue). **(E)** Maximal transcript abundances (NX) of the tissue of origin for all secretome proteins stratified according to annotated localization. Each protein is visualized as to whether it was detected (blue) or not detected (gray) in at least one of the assay platforms. **(F)** Number of blood secretome proteins detected by the three different methods.

**Proteins detected in the blood by antibody-based immunoassays**

We also compiled the plasma concentrations of proteins as detected by antibody-based immunoassays with a focus on the 730 proteins here defined as actively secreted to blood. A literature search was

performed for each protein, and references were collected for those studies that reported the absolute concentrations in human plasma. In total, reference plasma or serum concentrations were found for 365 of the 730 proteins secreted to the blood, and these values are presented on the Blood Atlas page of each gene and are combined

here for all of these proteins (Fig. 3B). The proteins are colored according to the functional annotation, and the results showed the expected high abundances for coagulation and complement factors, whereas cytokines, hormones, and growth factors often are present at low concentrations. Note that some of the actively secreted proteins with no annotated function were detected at relatively high concentrations in the blood. A comparison of the concentrations for 205 proteins determined by both proteomics and immunoassays showed a good correlation with a Spearman's  $\rho$  value of 0.79 (Fig. 3C).

### Proteins detected in the blood by PEAs

Last, we analyzed blood samples from a healthy cohort with a multiplex PEA (17), covering 748 proteins involving both secreted and leakage products to investigate the longitudinal variability of these proteins during 1 year. As an example, the abundances of the protein leptin were measured for 86 individuals across 1 year and four consecutive visits with 3-month intervals (Fig. 3D). Although the average protein abundances are greater in females, the leptin concentrations of individuals were highly variable, with several males having consistently higher amounts during the assay period. These data support an individual-based definition of reference values for diagnostic applications and suggest that caution should be taken when using population-based data as a reference for health. The protein abundances across the individuals during 1 year for all 748 proteins analyzed are presented as part of the new Blood Atlas.

We subsequently assessed the detectability of each of the 2641 secretome proteins based on the maximal transcript abundance in their respective tissue of origin (Fig. 3E). As expected, most of the proteins classified as blood proteins have been detected in blood, but it is also evident that many of the proteins classified in the other categories have been detected in blood, in particular many intracellular proteins and matrix proteins. The question arises as to whether this is due to these proteins having functional roles as circulating proteins or due to their leakage from the extracellular matrix. Furthermore, many of the proteins not detected by any assay were found to have low abundances in their predicted tissue of origin. On the basis of these three assay platforms, the overall detection of proteins secreted into the blood ( $n = 730$ ) was summarized (Fig. 3F). Note that 142 of the blood secretome proteins were not detected (or lacked specific assays) across all three platforms, and none of the assay platforms was able to detect more than half of the blood secretome proteins. This finding demonstrates the importance of developing assays for this important group of blood proteins and highlights the need for systematic efforts to develop assays for the “missing proteins” by respective assay platform.

### DISCUSSION

Here, we present an analysis of all of the proteins predicted to be secreted in humans based on sequence analysis of the corresponding transcripts, including all protein isoforms having a signal peptide and no transmembrane-spanning regions. We identified 2641 genes through this bioinformatics approach, and these genes were subsequently annotated individually to reflect the destinations of their products in the human body. The annotation provides a view of the actively secreted proteins in humans; however, the quality of the annotation varies across the different proteins, mainly due to lack of experimental evidence, in particular limited information regarding genes with multiple transcript isoforms. The annotations of the

individual proteins presented here will thus be revised when more information becomes available in the future. The results are presented in an updated version 19 of the HPA ([www.proteinatlas.org/blood](http://www.proteinatlas.org/blood)) to provide an open-access knowledge resource of the human secretome.

Some conclusions can be made from the results presented here. First, note that the number of predicted actively secreted blood proteins is surprisingly low ( $n = 730$ ), corresponding to less than 4% of all human protein-coding genes. Among the proteins identified in this category of the secretome are not only the classical plasma proteins, the inflammation proteins (cytokines and interleukins), well-known hormones, and receptors but also close to 100 proteins with no annotated function yet. Similarly, 88 proteins were annotated to be secreted to the gastric tract, which again is a low number, but the list includes not only well-known digestive enzymes and defense proteins but also proteins that are much less studied. The analysis also revealed that the largest number of proteins was annotated to remain locally after secretion, including many matrix proteins ( $n = 234$ ) and proteins that end up in specific tissues, including the brain and male and female tissues. However, the most unexpected finding from our analysis is that more than one-third ( $n = 932$ ) of the proteins are predicted not to be secreted out of the cell. This intracellular category reflects the fact that many proteins are retained in the ER, Golgi, or both or are sorted to various intracellular compartments. More in-depth studies are needed to annotate the functions of many of these proteins.

The proteins detected in human blood were also analyzed, and the study resulted in an annotated list of proteins that have thus far been detected by MS-based proteomics and antibody-based immunoassays. The analysis revealed that for a large fraction of the secreted blood proteins, assays are lacking and a quest for the future is therefore to extend the respective assays to cover all the actively secreted proteins defined here, including relevant leakage proteins, to provide a secretome-wide toolbox of assays for the proteins in the blood to expand our capabilities for functional analysis and precision medicine efforts. In conclusion, the open-access resource presented here includes a genome-wide classification of all protein-coding genes with regard to the predicted spatial location in the human body of the corresponding proteins, as well as showing the cellular and/or tissue origin of each of the secretome proteins to facilitate basic and applied research involving this important class of proteins.

### MATERIALS AND METHODS

#### Definition of the human secretome

The human secretome gene set includes Ensembl genes with either at least one splice variant encoding a protein with a signal peptide and no transmembrane regions according to HPA predictions (3) or for which UniProt has at least one isoform that is annotated as being secreted. From this gene set, the immunoglobulin-encoding genes were removed as were genes that encoded proteins for which there was little evidence of a secreted isoform.

#### Annotation of the human secretome

The genes were annotated into one of nine different categories based on the published literature, subcellular localization data from the HPA and UniProt, functional data from UniProt and HPA, RNA expression data from HPA, GTEx, and FANTOM, and protein expression data from HPA and UniProt. The annotation was

performed using an in-house annotation system displaying functional, expression, and subcellular location data from the different sources and links to relevant publications for each gene and the corresponding ensemble gene model with predicted signal peptides and transmembrane regions based on several algorithms (fig. S3). The annotation was made on a gene level, and for genes that were found to have both secreted and intracellular or membrane-bound protein isoforms, only the secreted variant was annotated. For each gene, a single category was selected together with an explanatory comment and references to relevant articles. A short description and the hierarchy of the categories are shown in table S1.

### Collection of protein plasma concentration references based on immunoassays

A literature search was performed for the 730 proteins annotated as being actively secreted to the blood, and the plasma concentrations of these proteins from up to three independent references were compiled. The search was limited to studies in which the protein concentrations had been determined using an antibody-based assay such as enzyme-linked immunosorbent assay, radioimmunoassay, or immunoturbidimetry. If reference concentrations from a mixed-sex healthy control group could be found, then these values were used, but if no such values were available, then concentrations from condition-specific groups were accepted, for example, from pregnant women or different diseases. In some studies, the plasma concentrations were not explicitly stated and median concentrations had to be estimated from plots in the articles. The complete list of collected plasma protein concentrations and their corresponding reference articles can be found in file S5.

### Retrieving MS-based plasma concentration data from the PeptideAtlas

For MS-based plasma proteomics, the data from >170 studies present in the PeptideAtlas ([www.peptideatlas.org](http://www.peptideatlas.org)) were used. We queried for the “Human Plasma Non-Glyco 2017-04” build and set presence levels “canonical,” “no redundant relationships” for redundancy, and “show estimated abundances” as display options. This search revealed 3694 entries, from which the 22 labeled as contaminants (“CONTAM”) were removed. Of the remaining 3672 entries, 3484 (93%) were listed with a protein concentration estimated by spectral counts column (“estimated\_ng\_per\_ml”) (18). These data were subsequently mapped to the ensemble gene set used in the HPA, and the 3223 remaining protein-coding genes were used in the analysis.

### Plasma profiling using PEA

Plasma from 86 healthy individuals being part of the Swedish SCAPIS SciLifeLab Wellness Profiling (S3WP) (21) project was used for this analysis. The S3WP study is an observational study with the aim to collect longitudinal data in a community-based cohort, and the subjects were recruited from the ongoing Swedish Cardio Pulmonary bioImage Study (SCAPIS) (22), which includes randomly selected subjects aged 50 to 65 years from the general Swedish population. Before sampling, all subjects fasted overnight, for at least 8 hours. Blood samples were drawn in EDTA tubes using standard operating procedures. All samples were stored at  $-80^{\circ}\text{C}$  until used. Plasma proteins were analyzed using a multiplex PEA (Olink Bioscience, Uppsala, Sweden). Each kit provides a microtiter plate for measuring 92 protein biomarkers in 90 samples (one panel), and in this study, 11 panels were used, including Cardiometabolic, Cell Regulation,

Cardiovascular II (CVD II), Cardiovascular III (CVD III), Development, Immune Response, Immuno-Oncology, Oncology II, Inflammation, Metabolism, Neurology, and Organ Damage. Each well in a kit contains 96 pairs of DNA-labeled antibody probes. Samples were incubated in the presence of proximity antibody pairs tagged with DNA reporter molecules. When the antibody pairs bind to their corresponding antigens, the corresponding DNA tails form an amplicon by proximity extension, which can be quantified by high-throughput, real-time polymerase chain reaction (PCR). Briefly, 1  $\mu\text{l}$  of each sample was mixed with 3  $\mu\text{l}$  of probe solution containing a set of 92 protein target-specific antibodies conjugated with distinctive DNA oligonucleotides. The mixture was incubated overnight at  $4^{\circ}\text{C}$ , and then 96  $\mu\text{l}$  of extension solution containing extension enzyme and PCR reagents was added. The generated fluorescent signal enabled the quantification of the protein using the BioMark HD System (Fluidigm Corporation). To minimize inter- and intra-run variation, the data were normalized using both an internal control (extension control) and an interplate control and were then transformed using a predetermined correction factor. The preprocessed data were provided in the arbitrary unit Normalized Protein eXpression (NPX) on a  $\log_2$  scale, which were then linearized by using the formula  $2\text{NPX}$ . Thus, a high NPX value corresponds to a high protein concentration. The limit of detection for each protein was defined as three SDs above the background. A list of the 748 analyzed proteins can be found in file S6.

### SUPPLEMENTARY MATERIALS

[stke.sciencemag.org/cgi/content/full/12/609/eaaz0274/DC1](http://stke.sciencemag.org/cgi/content/full/12/609/eaaz0274/DC1)

Fig. S1. Tissue of origin for proteins in different annotation categories.

Fig. S2. Expression profiles for the annotation categories.

Fig. S3. Some of the data displayed for the gene *LALBA* in the in-house annotation tool.

Table S1. Description of the annotation categories.

Data file S1. The immunoglobulin genes not annotated in this study.

Data file S2. The human secretome genes and their annotated categories.

Data file S3. Human blood secretome proteins that are targets for or act as FDA-approved pharmaceutical drugs according to DrugBank.

Data file S4. Proteins detected in blood by MS and their estimated concentrations.

Data file S5. Proteins detected in blood by immunoassays and the corresponding reference articles and concentrations.

Data file S6. Proteins detected by PEAs in a longitudinal study of 86 healthy individuals.

[View/request a protocol for this paper from Bio-protocol.](#)

### REFERENCES AND NOTES

1. M. Stastna, J. E. Van Eyk, Secreted proteins as a fundamental source for biomarker discovery. *Proteomics* **12**, 722–735 (2012).
2. H. F. Clark, A. L. Gurney, E. Abaya, K. Baker, D. Baldwin, J. Brush, J. Chen, B. Chow, C. Chui, C. Crowley, B. Currell, B. Deuel, P. Dowd, D. Eaton, J. Foster, C. Grimaldi, Q. Gu, P. E. Hass, S. Heldens, A. Huang, H. S. Kim, L. Klimowski, Y. Jin, S. Johnson, J. Lee, L. Lewis, D. Liao, M. Mark, E. Robbie, C. Sanchez, J. Schoenfeld, S. Seshagiri, L. Simmons, J. Singh, V. Smith, J. Stinson, A. Vagts, R. Vandlen, C. Watanabe, D. Wieand, K. Woods, M.-H. Xie, D. Yansura, S. Yi, G. Yu, J. Yuan, M. Zhang, Z. Zhang, A. Goddard, W. I. Wood, P. Godowski, A. Gray, The secreted protein discovery initiative (SPDI), a large-scale effort to identify novel human secreted and transmembrane proteins: A bioinformatics assessment. *Genome Res.* **13**, 2265–2270 (2003).
3. M. Uhlen, L. Fagerberg, B. M. Hallström, C. Lindskog, P. Oksvold, A. Mardinoglu, Å. Sivertsson, C. Kampf, E. Sjöstedt, A. Asplund, I. Olsson, K. Edlund, E. Lundberg, S. Navani, C. A. Szegedy, J. Odeberg, D. Djureinovic, J. O. Takanen, S. Hober, T. Alm, P.-H. Edqvist, H. Berling, H. Tegel, J. Mulder, J. Rockberg, P. Nilsson, J. M. Schwenk, M. Hamsten, K. von Feilitzen, M. Forsberg, L. Persson, F. Johansson, M. Zwaalen, G. von Heijne, J. Nielsen, F. Ponten, Tissue-based map of the human proteome. *Science* **347**, 1260419 (2015).
4. B. B. Sun, J. C. Maranville, J. E. Peters, D. Stacey, J. R. Staley, J. Blackshaw, S. Burgess, T. Jiang, E. Paige, P. Surendran, C. Oliver-Williams, M. A. Kamat, B. P. Prins, S. K. Wilcox, E. S. Zimmerman, A. Chi, N. Bansal, S. L. Spain, A. M. Wood, N. W. Morrell, J. R. Bradley, N. Janjic, D. J. Roberts, W. H. Ouwehand, J. A. Todd, N. Soranzo, K. Suhre, D. S. Paul, C. S. Fox, R. M. Plenge, J. Danesh, H. Runz, A. S. Butterworth, Genomic atlas of the human plasma proteome. *Nature* **558**, 73–79 (2018).



5. V. Emilsson, M. Ilkov, J. R. Lamb, N. Finkel, E. F. Gudmundsson, R. Pitts, H. Hoover, V. Gudmundsdottir, S. R. Horman, T. Aspelund, L. Shu, V. Trifonov, S. Sigurdsson, A. Manolescu, J. Zhu, O. Olafsson, J. Jakobsdottir, S. A. Lesley, J. To, J. Zhang, T. B. Harris, L. J. Launer, B. Zhang, G. Eiriksdottir, X. Yang, A. P. Orth, L. L. Jennings, V. Gudnason, Co-regulatory networks of human serum proteins link genetics to disease. *Science* **361**, 769–773 (2018).
6. A. A. Ellington, I. J. Kullo, K. R. Bailey, G. G. Klee, Antibody-based protein multiplex platforms: Technical and operational challenges. *Clin. Chem.* **56**, 186–193 (2010).
7. P. E. Geyer, N. A. Kulak, G. Pichler, L. M. Holdt, D. Teupser, M. Mann, Plasma proteome profiling to assess human health and disease. *Cell Syst.* **2**, 185–195 (2016).
8. N. J. Wewer Albrechtsen, P. E. Geyer, S. Doll, P. V. Treit, K. N. Bojsen-Moller, C. Martinussen, N. B. Jorgensen, S. S. Torekov, F. Meier, L. Niu, A. Santos, E. C. Keilhauer, J. J. Holst, S. Madsbad, M. Mann, Plasma proteome profiling reveals dynamics of inflammatory and lipid homeostasis markers after roux-En-Y gastric bypass surgery. *Cell Syst.* **7**, 601–612.e3 (2018).
9. D. R. Zerbino, P. Achuthan, W. Akanni, M. R. Amode, D. Barrell, J. Bhai, K. Billis, C. Cummins, A. Gall, C. G. Giron, L. Gil, L. Gordon, L. Haggerty, E. Haskell, T. Hourlier, O. G. Izuogu, S. H. Janacek, T. Juettemann, J. K. To, M. R. Laird, I. Lavidas, Z. Liu, J. E. Loveland, T. Maurel, W. McLaren, B. Moore, J. Mudge, D. N. Murphy, V. Newman, M. Nuhn, D. Ogeh, C. K. Ong, A. Parker, M. Patricio, H. S. Riat, H. Schuilenburg, D. Sheppard, H. Sparrow, K. Taylor, A. Thormann, A. Vullo, B. Walts, A. Zadissa, A. Frankish, S. E. Hunt, M. Kostadima, N. Langridge, F. J. Martin, M. Muffato, E. Perry, M. Ruffier, D. M. Staines, S. J. Trevanion, B. L. Aken, F. Cunningham, A. Yates, P. Flicek, Ensembl 2018. *Nucleic Acids Res.* **46**, D754–D761 (2018).
10. UniProt Consortium, UniProt: A worldwide hub of protein knowledge. *Nucleic Acids Res.* **47**, D506–D515 (2019).
11. H. R. Pelham, The retention signal for soluble proteins of the endoplasmic reticulum. *Trends Biochem. Sci.* **15**, 483–486 (1990).
12. J. E. Legakis, S. R. Terlecky, PTS2 protein import into mammalian peroxisomes. *Traffic* **2**, 252–260 (2001).
13. G. Schatz, The protein import system of mitochondria. *J. Biol. Chem.* **271**, 31763–31766 (1996).
14. P. Bornstein, Matricellular proteins: An overview. *J. Cell Commun. Signal.* **3**, 163–165 (2009).
15. D. S. Wishart, Y. D. Feunang, A. C. Guo, E. J. Lo, A. Marcu, J. R. Grant, T. Sajed, D. Johnson, C. Li, Z. Sayeeda, N. Assempour, I. Iynkkaran, Y. Liu, A. Maciejewski, N. Gale, A. Wilson, L. Chin, R. Cummings, D. Le, A. Pon, C. Knox, M. Wilson, DrugBank 5.0: A major update to the DrugBank database for 2018. *Nucleic Acids Res.* **46**, D1074–D1082 (2018).
16. F. Murtagh, P. Legendre, Ward's hierarchical agglomerative clustering method: Which algorithms implement Ward's criterion? *J. Classif.* **31**, 274–295 (2014).
17. E. Assarsson, M. Lundberg, G. Holmquist, J. Björkstén, S. B. Thorsen, D. Ekman, A. Eriksson, E. Rennel Dickens, S. Ohlsson, G. Edfeldt, A.-C. Andersson, P. Lindstedt, J. Stenvang, M. Gullberg, S. Fredriksson, Homogenous 96-plex PEA immunoassay exhibiting high sensitivity, specificity, and excellent scalability. *PLOS ONE* **9**, e95192 (2014).
18. T. Farrah, E. W. Deutsch, G. S. Omenn, D. S. Campbell, Z. Sun, J. A. Bletz, P. Mallick, J. E. Katz, J. Malmstrom, R. Ossola, J. D. Watts, B. Lin, H. Zhang, R. L. Moritz, R. Aebersold, A high-confidence human plasma proteome reference set with estimated concentrations in PeptideAtlas. *Mol. Cell. Proteomics* **10**, M110.006353 (2011).
19. J. M. Schwenk, G. S. Omenn, Z. Sun, D. S. Campbell, M. S. Baker, C. M. Overall, R. Aebersold, R. L. Moritz, E. W. Deutsch, The human plasma proteome draft of 2017: Building on the human plasma PeptideAtlas from mass spectrometry and complementary assays. *J. Proteome Res.* **16**, 4299–4310 (2017).
20. R. S. Tirumalai, K. C. Chan, D. A. Prieto, H. J. Issaq, T. P. Conrads, T. D. Veenstra, Characterization of the low molecular weight human serum proteome. *Mol. Cell. Proteomics* **2**, 1096–1103 (2003).
21. M. Neiman, C. Hellström, D. Just, C. Mattsson, L. Fagerberg, I. Schuppe-Koistinen, A. Gummesson, G. Bergstrom, O. Kallioniemi, A. Achour, R. Sallinen, M. Uhlén, P. Nilsson, Individual and stable autoantibody repertoires in healthy individuals. *Autoimmunity* **52**, 1–11 (2019).
22. G. Bergström, G. Berglund, A. Blomberg, J. Brandberg, G. Engström, J. Engvall, M. Eriksson, U. de Faire, A. Flinck, M. G. Hansson, B. Hedblad, O. Hjelmgren, C. Janson, T. Jernberg, A. Johnsson, L. Johansson, L. Lind, C. G. Löfdahl, O. Melander, C. J. Östgren, A. Persson, M. Persson, A. Sandstrom, C. Schmidt, S. Söderberg, J. Sundström, K. Toren, A. Waldenström, H. Wedel, J. Vikgren, B. Fagerberg, A. Rosengren, The Swedish CArdioPulmonary Biolmage study: Objectives and design. *J. Intern. Med.* **278**, 645–659 (2015).

**Acknowledgments:** We acknowledge the entire staff of the HPA program and the Science for Life Laboratory for their valuable contributions. **Funding:** Main funding was provided from the Knut and Alice Wallenberg Foundation (WCPR), the Erling Persson Foundation (KCAP), and the Novo Nordisk Foundation (CFB). **Author contributions:** M.U. conceived and designed the analysis. M.U., Å.S., M.J.K., A.H., B.F., F.E., F.P., J.N., L.S., J. Rockberg, J. Robinson, B.G.V., and D.K. collected and contributed data to the study. M.U., Å.S., W.Z., A.T., M.J.K., B.F., and L.F. performed the data analysis. M.U., Å.S., H.T., L.F., J. Rockberg, M.D., M.M., E.S., J.M., A.M., J.M.S., P.N., C.S., C.L., A.-L.V., M.L., A.B., A.-S.S., S.K., S.E., J.S., B.K., J.V., A.M., T.D.-C., E.P., R.S.H., M.-G.H., E. Lindström, B.A., D.M., D.S., P.T., F.D., and E. Lundberg performed the secretome annotation. K.v.F., P.O., and M.Z. provided the infrastructure for the data. G.B. and A.G. supplied clinical material. M.U. and Å.S. drafted the manuscript. M.U. and Å.S. revised the manuscript. All authors discussed the results and contributed to the final manuscript. **Competing interests:** The authors declare that they have no competing interests. **Data and materials availability:** All data needed to evaluate the conclusions in the paper are present in the paper or Supplementary Materials and are available without restriction at the HPA portal ([www.proteinatlas.org](http://www.proteinatlas.org)).

Submitted 6 August 2019  
Accepted 12 November 2019  
Published 26 November 2019  
10.1126/scisignal.aaz0274

**Citation:** M. Uhlén, M. J. Karlsson, A. Hober, A.-S. Svensson, J. Scheffel, D. Kotol, W. Zhong, A. Tebani, L. Strandberg, F. Edfors, E. Sjöstedt, J. Mulder, A. Mardinoglu, A. Berling, S. Ekblad, M. Dannemeyer, S. Kanje, J. Rockberg, M. Lundqvist, M. Malm, A.-L. Volk, P. Nilsson, A. Månberg, T. Dodig-Crnkovic, E. Pin, M. Zwahlen, P. Oksvold, K. von Feilitzen, R. S. Häussler, M.-G. Hong, C. Lindskog, F. Ponten, B. Katona, J. Vuu, E. Lindström, J. Nielsen, J. Robinson, B. Ayoglu, D. Mahdessian, D. Sullivan, P. Thul, F. Danielsson, C. Stadler, E. Lundberg, G. Bergström, A. Gummesson, B. G. Voldborg, H. Tegel, S. Hober, B. Forsström, J. M. Schwenk, L. Fagerberg, Å. Sivertsson, The human secretome. *Sci. Signal.* **12**, eaaz0274 (2019).

## The human secretome

Mathias Uhlén, Max J. Karlsson, Andreas Hober, Anne-Sophie Svensson, Julia Scheffel, David Kotol, Wen Zhong, Abdellah Tebani, Linnéa Strandberg, Fredrik Edfors, Evelina Sjöstedt, Jan Mulder, Adil Mardinoglu, Anna Berling, Siri Ekblad, Melanie Dannemeyer, Sara Kanje, Johan Rockberg, Magnus Lundqvist, Magdalena Malm, Anna-Luisa Volk, Peter Nilsson, Anna Månberg, Tea Dodig-Crnkovic, Elisa Pin, Martin Zwahlen, Per Oksvold, Kalle von Feilitzen, Ragna S. Häussler, Mun-Gwan Hong, Cecilia Lindskog, Fredrik Ponten, Borbala Katona, Jimmy Vuu, Emil Lindström, Jens Nielsen, Jonathan Robinson, Burcu Ayoglu, Diana Mahdessian, Devin Sullivan, Peter Thul, Frida Danielsson, Charlotte Stadler, Emma Lundberg, Göran Bergström, Anders Gummesson, Björn G. Voldborg, Hanna Tegel, Sophia Hober, Björn Forsström, Jochen M. Schwenk, Linn Fagerberg and Åsa Sivertsson

*Sci. Signal.* **12** (609), eaaz0274.  
DOI: 10.1126/scisignal.aaz0274

### Secrets of secreted proteins

Although proteins that are actively secreted from cells constitute a large proportion of the targets of pharmaceutical drugs, we lack a complete understanding of the size and composition of this group of proteins, collectively referred to as the human "secretome." Uhlén *et al.* analyzed the genes of the Human Protein Atlas that encode proteins that are predicted to be secreted and combined this with data derived from antibody- and mass spectrometry-based analyses of the relative abundances, sites of origin, and final locations of proteins within the body. Their analysis suggests that many proteins that were thought to be secreted are retained intracellularly, and that another group of secreted proteins are retained locally rather than being released into the blood. These data serve as a resource for future research on this important class of proteins.

#### ARTICLE TOOLS

<http://stke.sciencemag.org/content/12/609/eaaz0274>

#### SUPPLEMENTARY MATERIALS

<http://stke.sciencemag.org/content/suppl/2019/11/22/12.609.eaaz0274.DC1>

#### RELATED CONTENT

<http://stke.sciencemag.org/content/sigtrans/12/565/eaaw3483.full>  
<http://science.sciencemag.org/content/sci/357/6352/eaan2507.full>  
<http://science.sciencemag.org/content/sci/356/6340/eaal3321.full>

#### REFERENCES

This article cites 22 articles, 7 of which you can access for free  
<http://stke.sciencemag.org/content/12/609/eaaz0274#BIBL>

#### PERMISSIONS

<http://www.sciencemag.org/help/reprints-and-permissions>

Use of this article is subject to the [Terms of Service](#)

*Science Signaling* (ISSN 1937-9145) is published by the American Association for the Advancement of Science, 1200 New York Avenue NW, Washington, DC 20005. The title *Science Signaling* is a registered trademark of AAAS.

Copyright © 2019 The Authors, some rights reserved; exclusive licensee American Association for the Advancement of Science. No claim to original U.S. Government Works